DOCUMENT RESUME

ED 454 601 EA 031 089

AUTHOR Achilles, C. M.; Finn, J. D.; Gerber, Susan B.
TITLE Small Classes Do Reduce the Test-Score Achievement Gap.
SPONS AGENCY Spencer Foundation, Chicago, IL.
PUB DATE 2000-10-26
NOTE 53p.; Paper presented at the Annual Meeting of the Council of Great City Schools (Los Angeles, CA, October 26, 2000). Written with H. Pate-Bain, J. Boyd-Zaharias, and G. M. Pannozzo.
PUB TYPE Reports - Descriptive (141) -- Speeches/Meeting Papers (150)
EDRS PRICE EDRS Price MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS *Academic Achievement; *Achievement Gains; *Class Size; Primary Education
IDENTIFIERS *Student Teacher Achievement Ratio Project TN

ABSTRACT

Tennessee's Project STAR, a randomized experiment involving almost 12,000 pupils, demonstrated convincingly that small classes in the early elementary (K-3) grades increase pupil performance, reduce the test-score achievement gap between or among different social groups, and can have long-lasting effects. The benefits are greater for minority students, resulting in a reduced minority-White achievement gap. This paper summarizes latest research on the immediate and long-term effects of small classes. Section 1 discusses the need for independent research on important education issues such as class size and achievement gap. Section 2 begins with a call for clear definition of terms in any class-size discussion. Section 3 presents evidence from recent non-STAR sources related to class size and student achievement. Section 4 offers a summary, conclusions, and some recommendations, including conditions for implementing small classes, problems that districts have encountered, and conditions likely to maximize small-class benefits. Three figures and 14 tables show trends in data. Appendices include a description of STAR, a summary of class-size findings and issues derived from various sources, part of Texas HB 72 which mandated small classes, a major caveat on Texas data, and a discussion about the Black-White test score gap. (RT)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 454 601

# SMALL CLASSES DO REDUCE THE TEST-SCORE ACHIEVEMENT GAP

*Tennessee's Project STAR, a randomized experiment involving almost 12,000 pupils, demonstrated convincingly that small classes in the early grades increase pupil performance. The benefits are greater for minority students, resulting in a reduced minority—White achievement gap. This presentation will summarize the latest research on the immediate and long-term effects of small classes. It will also discuss conditions for implementing small classes—the problems that districts have encountered and conditions likely to maximize small-class benefits.*

A Paper Presented at the Council of Great City Schools Conference

C. M. Achilles, EdD
Professor, Education Leadership
304 Porter Ed. Building
Eastern Michigan University
Ypsilanti, MI, 48197

J. D. Finn, PhD
Professor, Graduate School of Education
Christopher Baldy Hall
SUNY, Buffalo, NY 14260

Susan B. Gerber, PhD
Technology Center
Christopher Baldy Hall
SUNY, Buffalo, NY 14260

Gina Pannozzo
Graduate Assistant
Christopher Baldy Hall
SUNY, Buffalo, NY 14260

Jayne Boyd-Zaharias
HEROS, Inc.
Lebanon, TN 37087

October 26, 2000

Los Angeles, CA

BEST COPY AVAILABLE

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY

C.M. Achilles

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)



EA 031089

# Small Classes Do Reduce The Test Score Achievement Gap *

There continues to be considerable concern about closing the "achievement gap." This usually means trying to make more equal the average group scores on some measure, such as a standardized or normed test, between "majority" and "minority" students, often defined by race or ethnicity (Hispanic, African-American, Anglo, Native American, Asian-American, etc.). There has also been concern for the achievement gap between males and females, for example, on math (favoring males) or in early grades on language arts (favoring females). Title I is emphasizes closing the achievement gap between "disadvantaged" and not-disadvantaged students (etc.). The concern for achievement gaps in education outcomes between and among groups is not exclusively a concern of U. S. educators. Progress in solving this conundrum will improve education experiences and social benefits.

This paper offers evidence that small classes in early elementary grades (K-3) reduces the test-score achievement gap between or among groups, such as Anglo and Asian American students and African-American and Hispanic students. The paper presents this argument in four sections. Section 1 discusses the need for independent research on important education issues such as class size and achievement gap. Section 2 which provides data from Project STAR (Student Teacher Achievement Ratio), a longitudinal, large-scale randomized educational experiment, begins with a call for clear definition of terms in any class-size discussion. Section 3 presents direct and indirect evidence from recent non-STAR sources related to class size and

---

* C. M. Achilles, J. D. Finn, and S. Gerber are primary authors. They have been assisted by H. Pate-Bain, J. Boyd-Zaharias, and G. M. Pannozzo, and by the research of others cited throughout the paper. This "work in progress" is partially supported by a grant from The Spencer Foundation, "A Study of Class Size and At Risk Students." Some tables, figures, etc. are "cut and paste" from extant documents and may be in different formats, fonts, etc.

student achievement. Section 4 offers a summary, conclusions, and some recommendations. Primary research, news media, state and district statistics, and the professional literature, are all used in this presentation.

## Section 1. Independent Research.

In election years there is much rhetoric about "fixing" the schools. Yet, as a recent NY Times article explained, folks are skeptical because the recommended "quick fixes" are often ideological and high-sounding, but seldom based on professional knowledge of schooling. Or, the remedies are proffered based upon the research of persons or groups who hardly could be called "disinterested." (Advocacy, rather than research).

Archibold (2000) quoted a school superintendent who noted that he could walk into any store in his community and no-one he spoke to would have any hesitation about telling him how to run the school system (probably the largest business in town!). "The overriding problem, as he sees it, is a dearth of independent research on what actually works. The slightest whiff of apparent success--a limited study... yields a headlong, and possibly, misplaced rush of support, he said." (p. v27, Emphasis Added).

Recently the Phi Delta Kappan featured the extension of a debate that was earlier carried by the less-read Educational Researcher. The debate was whether or not "Success For All" (SFA) actually provided success for students (Pogrow, 2000 a, b; Slavin, 2000; Slavin & Madden, 2000). Pogrow, the developer and advocate for Higher Order Thinking Skills (HOTS) questioned both the research and evaluation outcomes for SFA and whether the evaluation work--all done by or commissioned through SFA personnel--could be considered independent and/or disinterested. Slavin (2000) and Slavin and Madden (2000) summarized findings that SFA was, indeed, successful. Pogrow (2000 b) countered that Slavin and Madden still relied on data from

persons closely connected to SFA. Butzin (2000) helps focus on the question of "disinterested": "And I am certainly envious that (SFA) and Robert Slavin's research center have managed to capture $56.14 million in federal grants, plus generating $44 million in program sales" (p. 98). Indeed, SFA is "certified" as a Comprehensive School Reform Model, and is strongly recommended to schools as a program to receive Title I funds. Achilles (1998) questioned if SFA's results were equal to class-size results and if SFA got its outcomes because of small classes. Most school "remedial" projects are expensive. Some were developed with tax dollars at federal R & D centers, labs, and sponsored-research sites. Levin (1998) raised a question about the cost to local school systems of such federally-developed programs:

> In fact, I believe that they ought to be made available at low cost and question why they are sold commercially at high prices, given the substantial grants and subsides provided for their development by government agencies. . . presumably to reduce costs and promote their use (Levin, 1998, p. 9).

Of substantial interest in this discussion of independent research is the question of the initiation of, support for, and independent nature of the class-size research relied upon in this paper. Much support for small classes and for this discussion of a test-score achievement-gap reduction comes from the Tennessee class-size experiment, Project STAR conducted in 1984-1989, and from the extension of STAR research 1989-2000 (and beyond) through the Lasting Benefits Study (LB S), Project Challenge, and various STAR follow-up studies. (Table 1 summarizes STAR and STAR-related studies (1984-2000) used in this discussion. Ritter and Boruch (1999) reviewed STAR's initiation and processes and makes it clear that STAR was conducted as an independent study to answer class-size questions.

Table 1. Samples of the STAR Legacy of Class-size Studies, Categorized as "Subsidiary" (directly from STAR), "Ancillary" (building on the STAR database) and "Related" (usually involving STAR researchers and using STAR information and earlier findings).

| CATEGORY, TITLE & PURPOSE * | DATE(S) | AUTHOR(S), SOURCE, DATE |
|---|---|---|
| STAR Pilot (DuPont) | 1984-1986 | Bain et al. |
| STAR (Class-size experiment) | 1985-1989 | Word et al., 1990 Finn & Achilles, 1990 |
| Subsidiary Studies | | |
| • Lasting Benefits Study (LBS) | 1989-1996 | Nye et al., 1991-1999 |
| • Project Challenge (TN) | 1989-1996 | Nye et al., 1991-1995 |
| • Participation, Grades 4, 8 | 1990, 1996 | Finn, 1989, 1993; Voelkl, 1995; Finn et al., 1989; Finn & Cox, 1992 |
| • STAR Follow-up Studies | 1996-2000 | HEROS, 1997-2000 |
| Ancillary Studies | | |
| • Retention in Grade | 1994-1995 | Harvey, 1994, 1995 |
| • Achievement Gap | 1993-1995 | Bingham, 1993, 1994; Achilles et al., 1997-98 |
| • Value of K in Classes of Varying Sizes (test scores) | 1985-1989 | Achilles, Bain, Nye, 1994 |
| • School Size and Class-Size Issues | 1985-1989 | Nye, K., 1995 |
| • Random v. Non-Random Pupil Assignment and Achievement | 1985-1989 | Zaharias et al., 1995 |
| • Re-analysis, Sample "drift" (out-of-range classes) | 1985-1989 | Boyd-Zaharias et al., 1995 Finn et al., 1999 |
| • Class Size and Discipline Grades 3,5,7 | 1989, 1991, etc. 1996 | Several studies. SSS, 1995; Hibbs (1997). |
| • Outstanding Teacher Analysis (top 10% of STAR teachers) | 1985-1989 | Bain et al., 1992 |
| • Teacher Aides (continuing) | | Achilles et al., 1994; Gerber; Zaharias & Bain, 1998 |
| • Continuing student growth | 1985-2000 | Finn, Achilles et al.; Bain et al. |
| • College entrance exams | 2000 | Krueger & Whitmore |
| Related Studies | | |
| • Success Starts Small (SSS): A Study in 1:14 and 1:23 Schools | 1993-1995 | Achilles et al., 1994 Kiser-Kling, 1995 |
| • Burke Co., NC Study | 1992-2000 | Achilles et al., 1995 SERVE, 1996 ; Harman et al., 1998 |
| • Education Production Functions | 1996-2000 | Krueger, 1997, 1998, 1999, 2000 |

* This is a sample of STAR-related class-size studies. Not all authors appear exactly as listed here. A similar table appears in several STAR reports and articles. Others have reported on STAR (e.g., Mosteller, 1995; Finn, 1998; McRobbie et al., 1998); several have conducted secondary re-analyses of STAR data (e.g., Goldstein & Blatchford, 1998; Krueger, 1997-2000).

Project STAR was initiated in 1984 by legislation. The Tennessee Legislature needed data upon which to base legislation on elementary-grade class sizes. A team of four researchers representing major Tennessee universities (U. TN; Vanderbilt; U of Memphis; TN State University), consultants, and advisory groups representing superintendents, legislators, researchers, and state education officials conducted the experiment to determine the effect of small classes (about 15 students per teacher) on the achievement and development of primary-grade youngsters (grades K-3). STAR was funded by approximately $14 million (1985-1990), appropriated through the legislature, and by modest annual supplements to support LBS and Follow-up Studies (1990-2000). Analyses have been supported by grants from the Spencer Foundation (e.g., A Study of Class Size and At-Risk Students). Other sources of data include SAGE (Student Achievement Guarantee In Education) in Wisconsin, a state-funded initiative; Success Starts Small or SSS, an observation study of teaching in small and regular classes supported by a grant from the state of North Carolina, and inferences based on data from selected states where state or system-wide class-size reductions (CSR) have been initiated such as Texas (House Bill 72 in 1984) and California (1996). More than half of the states have enacted some class-size efforts by 2000.

As a check on the independent nature of STAR stud, we compared STAR processes and outcomes with seven criteria that have been suggested to assess whether or not school or social programs "work." The comparisons in Table 2 show that STAR meets or exceeds each criterion. In fact, most actual class-size research is independent, as it is designed open-endedly to answer questions about class size and its influence on student outcomes. The further that the studies and analyses move from a true experimental design, the more likely that the design and methods are less rigorous, and that more care is required in interpreting and generalizing results. The class-

size research used here qualifies as "independent." There is one caveat. Although one tries to maintain objectivity, years of work (1984-2000) on topics as basic to education outcomes as the size of the group that the teacher works with daily may push the researcher a bit toward advocacy, especially when well-constructed class-size research results consistently support the use of small classes in the early years of schooling for short and long-term gains.

---

Table 2. Comparison of STAR results with Crane's Seven Criteria for Social Programs that Work. STAR is a class-size reduction (CSR) experiment, not a pupil teacher ratio (PTR) effort. *

| CRANE CRITERIA and QUESTIONS | STAR'S FACTS |
|---|---|
| 1. Do the benefits outweigh the costs? | 1. In the short term (K-3), there were no definitive data. In the "follow-up studies;" yes; in the STAR reanalysis, yes; in alternative implementations, yes. See Krueger (1997, 1998; Finn & Achilles, 1999). |
| 2. Does the program have a statistically significant effect on the treatment group? | 2. Yes. This statistically significant difference was found each year, all years, and in many combinations of analyses done by STAR persons and by others (as far away as London). |
| 3. What is the magnitude of the program's effect? | 3. Although some effect-size (ES) results were .17-.40 in the early analyses, these are being re-computed. Effects will be larger overall and about twice as high for minority children as for Anglo children, grades K-3 (each year, all years). Grade-equivalent analyses show continuing growth even after students leave small classes (see #5). |
| 4. How long do the effects of the program last? | 4. HEROS Inc. and Krueger are conducting longitudinal analyses into H.S. and beyond. Positive academic and social effects of K-3 small classes are highly visible even in H.S. and beyond. |
| 5. What is the relationship of the evaluator to the program.? | 5. The original evaluator of STAR was a contracted independent design and analysis expert. STAR personnel did secondary and individual analyses, but the external expert's work is (and was) the primary analysis accepted and published. |
| 6. Can the program and its results be replicated?<br>7. Can the program maintain its effectiveness on a larger scale? | 6. & 7. They have been consistently replicated in well designed class-size analyses, and even in large-scale, statewide efforts. Almost exact replications of STAR have been achieved in single districts, in states, and in general policy implementations. The reported gains and effect sizes for the various well conducted studies are all similar. Evaluations in CA, and the results in Texas (HB 72, 1984) suggest large-scale benefits. |

* Social Programs That Work edited by Jonathan Crane (1998). Russell Sage Foundation, 111 East 64th Street, New York, New York 10021. 324 pages. Reported by Wagner, C. in FUTURIST, (1999, April), p. 18.

## Section II. Data and Conclusions from STAR and STAR-Related Studies.

### The Definitional Issue

At the outset of a class size discussion, there needs to be a clear definition of class size. Class size is the number of youngsters in the teacher's classroom and for whom that teacher is responsible and accountable. Class size is determined by adding the students in the classroom or by establishing the number of students in the room through an assignment process. Pupil-Teacher Ratio (PTR), a term often incorrectly substituted for class size, is much more murky than class size.

A ratio such as PTR is calculated by division, often using aggregated data for students. Akerhielm (1995), an economist, explained PTR/class size differences in detail (p.230).

> In addition, pupil-teacher ratios are defined as the number of students in the school divided by the number of full-time teachers for an entire school and may have nothing to do with actual class size. This ratio often includes guidance counselors, principals, and special education teachers in the count of teachers, and thus the lower the ratio, the higher the non-teaching staff, regardless of actual class size. Moreover, even if the ratio only includes actual teachers, schools with the same pupil-teacher ratio may have significantly different class sizes . . ..
>
> If a school has a deliberate policy to assign difficult or less able students to smaller classes, then any positive effect of small class size on student performance may be disguised because such students may tend to score lower on tests.

Boozer and Rouse (1995) described PTR and class size differences and related these to achievement. The correlations they found between PTR and class size were weak, ranging from .13 to .26.

Boozer and Rouse's findings help explain why PTR changes do not influence student outcomes much. "Once again we find that the pupil teacher ratio does not (statistically) increase

in schools with a larger proportion of black students, but that the average class size does." (p. 8) and . . . "On the other hand, students in schools with larger average class sizes have significantly smaller test-score gains" (p. 8). Boozer and Rouse (1995) find another important class-size difference from PTR. "Thus, using only the cross-school variation in class size we find that it satisfies two requirement for explaining the black-white gap in achievement: black students are in larger classes than are white students, and smaller classes are associated with larger test score gain." (p. 9). Left unsaid is that PTR change is most often influenced by remedial efforts and "projects" to help at-risk students, projects like Title I, the nation's largest "remedial education" effort. Remedial projects and special education efforts seldom produce large test-score gains: Neither do Title I efforts (Abt Associates, 1997; Borman & D'Agostino, 1996; Wong & Meyer, 1998). But, they require teachers and space.

Class size and PTR are two different constructs and they should not be confused. They generate different student outcomes and should be used differently. Table 3 compares Class Size (CS) and of PTR on several dimensions, showing how PTR and CS differ. The remainder of this paper focuses on class-size and CSR outcomes and to the extent practicable, the discussion is not about PTR outcomes. Note that SFA is really a project and is much more like PTR than class size. So is Reading Recovery (RR) which is a tutorial with specially-trained teachers.

Table 3. Some Major Differences Between Class Size (CS) and Pupil-Teacher Ratio (PTR).

| VARIABLES of note in comparing PTR and Class Size | PUPIL-TEACHER RATIO (PTR) | CLASS SIZE (CS) or (CSR) for Class-Size Reduction |
|---|---|---|
| Definition | The number of students at a site (building, district, class) divided by: teachers, educators, adults, (etc.) serving the site. | The number of students in a teacher's room regularly for whom that teacher is responsible and accountable. |
| Computation | DIVISION, with various divisors available depending upon the EXACT definition. | ADDITION. CS cannot be accurately determined from large databases. |
| Concept | The teacher needs help; the student needs special services the teacher cannot provide. | A competent teacher can handle most education issues if given a reasonable case load. |
| Operation and Context | A project and "pull-out"-driven model full of commotion and "Band Aid" treatments. Loss of time on task. Difficulty in determining responsibility and accountability. | Teacher is responsible for the student's growth and development: (ABCD): Academics, Behavior, Citizenship/participation, Development. Small focused learning groups. |
| Outcomes | CONSISTENTLY MARGINAL. Note, for example, education "production function" analyses; Title I evaluations, Boozer and Rouse (1995), Borman and D'Agostino (1996) Wong and Meyer (1998), etc. | CONSISTENTLY POSITIVE on many variables (A, B, C, D). Many studies (1970-2001) such as SAGE, Fairfax County, Glass and Smith, Burke County, Krueger. (See also Table 1). |

Project STAR is a True Education Experiment

Details of the STAR experiment have been extensively reported in research and class-size literature. [See, for example, Finn and Achilles (1990, 1999), Mosteller (1995), Word et al., (1990), Ritter and Boruch (1999), Finn (1998), Achilles (1999)]. A synopsis of the major elements of the STAR experiment is presented in Appendix A taken from Achilles and Finn, (1999). The size, scope, and careful attention to scientific rigor, longitudinal nature of the follow-up data, and the various confirming studies mark STAR as an important study in education, and an important source of information about education effects. Appendix B offers a synopsis of key class-size findings from STAR and related studies. Professor Mosteller (emeritus) of Harvard said of STAR (1995):

> . . . the Tennessee class size project, a controlled experiment which is one of the most important investigations ever carried out and illustrates the kind and magnitude of research needed in the field of education to strengthen schools (p. 113).
>
> . . . the Tennessee class size project, . . . illustrates the kind and magnitude of research needed in the field of education to strengthen schools (p. 113). . . . it is important that both educators and policy makers have access to its statistical information and understand its implications. (p. 126).
>
> Project STAR (Student/Teacher Achievement Ration), a study of the educational effects of class size in the state of Tennessee (Word et al., 1990) is one of the great experiments in education in U.S. history. (Mosteller, Light & Sachs, 1996, p.814).

The Achievement Gap Results

Results of achievement-gap analyses of STAR and STAR-related studies are presented in two parts. Part 1 includes data directly from the K-3 outcomes, relying mostly on test scores. Part 2 incorporates information from the STAR follow-up work, including high school graduation rates, college-entrance test-taking data, and general "Big adult" outcomes.

Part 1. In a commissioned paper for the Department of Education, Finn (1998) provided the effect sizes for STAR small classes, grades K-3, for white and for minority students. The following material is taken directly from Finn (1998, pp. 9-10).

> In every instance, small classes outperformed the other class types; effect sizes for the total sample *(All)* range from about .15σ in kindergarten to about .25σ in grades 1, 2, and 3.[9] And like the research that preceded STAR, the small-class advantage was consistently greater for minority students (most of whom were black) than for whites. In most comparisons, the impact on minorities was about twice as large as it was for white students. This resulted in a considerably reduced achievement gap. In reporting this effect, Finn and Achilles (1990) noted that the difference between minorities and whites in mastery rates on the grade 1 reading test was "reduced from 14.3 percent in regular classes to 4.1 percent in small classes" (p. 568).
>
> Two additional points should be noted. First, the effect sizes in table 1 show that small classes present up to a 1/4σ advantage compared to larger classes in every subject tested. [10] Although the researchers did not devise methods for computing the total impact on achievement, it is greater than any single difference would indicate. Second, the effect sizes in table 1 actually underestimate the true small-class advantage. An unavoidable phenomenon during the 4-year project was the "drifting" of some classes out of the target size range, as students transferred into or out of a class or school. Preliminary indications are that the effect sizes would be substantially greater if out-of-range classes were removed from the data. [11]
>
> In sum, due to the magnitude of the Project STAR longitudinal experiment, the design, and the care with which it was executed, the results are clear:
>
> - This research leaves no doubt that small classes have an advantage over larger classes in student performance in the early primary grades.
>
> At the same time, the research leaves behind a wealth of data that have only begun to be analyzed for what they can tell us.

---

[9] Although precise grade equivalents are not available, these differences correspond to an advantage of about .1 grade equivalents (or about I month) by the end of kindergarten, about .2 grade equivalents (or about 2 months) at the end of first grade, and somewhat more by the end of grade 2.

[10] Including several other subtests not listed in table 1.

[11] In the range .3σ and upward (Zaharias, *et al.*, 1995).

[12] Each year (1990—1994) the number of students tested was between approximately 4,200 and 6,000.

Table 4. (Table 1 Reprinted from Finn, 1998, p.10)

Table 1. Small-class effect sizes, grades kindergarten (K) through 3, by skills, motivation, and self-concept data.

| Scale | Group | Grade Level | | | |
|---|---|---|---|---|---|
| | | K | 1 | 2 | 3 |
| Word Study Skills | W | .15 | .16 | .11 | N/A |
| | M | .17 | .32 | .34 | |
| | **ALL** | **.15** | **.22** | **.20** | |
| Reading | **W** | .15 | .16 | .11 | .16[a] |
| | M | .15 | .35 | .26 | .35[a] |
| | **ALL** | **.18** | **.22** | **.19** | **.25**[a] |
| Total Reading | W | -- | .17 | .13 | .17 |
| | M | -- | .37 | .33 | .40 |
| | **ALL** | **.18** | **.24** | **.23** | **.26** |
| Basic Skills | W | | 4.8% | 1.6% | 4.0% |
| First (BSF) | M | N/A | 17.3% | 12.7% | 9.3% |
| Reading | **ALL** | | **9.6%** | **6.9%** | **7.2%** |
| Total | W | .17 | .22 | .12 | .16 |
| Mathematics | M | .08 | .31 | .35 | .30 |
| | **ALL** | **.15** | **.27** | **.20** | **.23** |
| Basic Skills | W | | 3.1% | 1.2% | 4.4% |
| First (BSF) | M | N/A | 7.0% | 9.9% | 8.3% |
| **Mathematics** | **ALL** | | **5.9%** | **4.7%** | **6.7%** |
| Motivation | W | .00 | -.02 | -.03 | -.01 |
| | M | .03 | -.01 | .07 | .11 |
| | **ALL** | **.01** | **.00** | **.01** | **.00** |
| Self-Concept | W | .10 | .07 | .00 | -.05 |
| | M | .10 | .05 | .03 | .04 |
| | **ALL** | **.11** | **.07** | **.02** | **.02** |

NOTE: The values for BSF Reading and BSF Mathematics represent differences in the percent passing (no standard deviation). All other values are mean differences: Small - (Regular + Aide)/2, divided by the standard deviation of the scale. Standard deviations computed for all students in regular classes, and all white (W) and minority (M) students separately.

[a] Total Language scale in grade 3 (not Reading).

Because criterion-referenced tests (CRT) measure most closely the content that is taught in schools, the advantages of small classes in reducing achievement gaps is presented here using CRTs. Table 5 shows the overall differences in the percent correct on Tennessee's Basic Skills First (BSF) Tests by grades, by class types of Small (S), Regular (R), and Regular/Aide (RA). The BSF is a CRT tied directly to the state's learning objectives. This analysis does not consider duration—how long a student spent in a class type.

Table 5. Differences Between Minority (M) and Non-minority (NM) and Free Lunch (FL) and Not Free Lunch (NFL) Students by Class Type (S, R, RA) on BSF Percent Correct.

| | | READING BSF (%) | | | MATH BSF (%) | | |
|---|---|---|---|---|---|---|---|
| Difference | Grade | S | R | RA | S | R | RA |
| (M) (NM) | 1 | 2.6 | 10.6 | 8.9 | 5.4 | 9.8 | 8.9 |
| (FL) (NFL) | 1 | 9.1 | 14.0 | 10.3 | 7.5 | 10.1 | 9.0 |
| (M) (NM | 2 | 3.8 | 6.1 | 6.4 | 5.0 | 7.7 | 8.0 |
| (FL) (NFL) | 2 | 4.3 | 6.6 | 5.8 | 5.6 | 8.3 | 7.5 |
| (M) (NM | 3 | 10.1 | 11.9 | 10.6 | 11.6 | 13.5 | 12.4 |
| (FL) (NFL) | 3 | 12.7 | 11.8 | 12.7 | 10.1 | 10.4 | 12.1 |

This comparison shows that in only one of 24 comparisons of (S) with (R) or with (RA) (grade 3, S with R for FL and NFL) does the small class (S) NOT provide the smallest, or tie for smallest (n=1) percentage difference between the two achievement groups compared (minority with non-minority, and free lunch with no free lunch. There is a general trend toward less achievement-gap difference as the scores are reported for later grades. As the next analysis shows, this decline in gap reduction is at least partially attributed to the length of time that a student is in the same class type (S), (R), or (RA). Differences in the amount of time a student

spent in a class type are related to several factors, including mobility, retention in grade, and a re-randomizing of students in the (R) and (RA) conditions after Kindergarten (K) when no differences were found between the test-score results of (R) and (RA) students. Tables 6 and 7 show the differences in the percent of objectives mastered (correctly answering at least 3 of 4 questions for each objective) by class type (S), (R), (RA), by number of years a student was in that class type at the time of testing. For example, at grade-1 testing a student could have been in a (S) class for 1 year (grade 1 only) or for 2 years (K and 1). Table 6 presents results for the minority (M) and non-minority (NM) gap; Table 7 shows the free lunch (FL) and not free lunch (NFL) results. The smaller the percent difference, the less "achievement gap" there is. Tables 6 and 7 provide clear evidence that the small-class condition offers a substantial achievement-gap reduction for race/ethnic and for SES differences.

In all cases reported in Tables 6 and 7 the percentages (scores) for (S) are considerably higher than for (R) and (RA). These differences were always statistically significant at $p \leq .01$.

[TABLES 6 and 7 ABOUT HERE]

Retention in grade usually falls heaviest on minority, low SES, and male students. To the degree that class-size influences retention in grade, it can be said to reduce the equity/achievement gap in schooling. In all cases, grades K-3, the (S) condition has the lowest percentage of students retained. Is this cause of effect? Table 8 provides a summary of the retention data in STAR for K-3. (Word et al., 1990, p. 170). [TABLE 8 ABOUT HERE]

Table 6. Summary of Differences Between Minority (M) and Non-Minority (MN) Students by Class Type (S, R, RA) on Percent of BSF Objectives Mastered by Years in a Class Type at Grades 1,2, and 3. (Percents Rounded).

| | | READING BSF (%) | | | MATH BSF (%) | | |
|---|---|---|---|---|---|---|---|
| | | S | R | RA | S | R | RA |
| Grade 1 | 1 yr (1) | 6 | 8 | 8 | 7 | 9 | 9 |
| | 2 yrs (K,1) | 2* | 12 | 9 | 5 | 10 | 8 |
| Grade 2 | 1 yr (2) | 6 | 16 | 15 | 3 | 8 | 6 |
| | 2 yrs (1, 2) | 9 | 15 | 12 | 4 | 6 | 7 |
| | 3 yrs (K, 1, 2) | 7 | 11 | 14 | 5 | 8 | 8 |
| Grade 3 | 1 yr (3) | 3 | 13 | 8 | 7 | 14 | 11 |
| | 2 yrs (2, 3) | 9 | 14 | 8 | 10 | 16 | 14 |
| | 3 yrs (1, 2, 3) | 13** | 8 | 7 | 17** | 7 | 7 |
| | 4 yrs (K, 1, 2, 3) | 6 | 8 | 9 | 9 | 10 | 12 |

* Note that the "gap" has not opened (2%) here enough to be significant.

** In only two places (grade 3, 3 years) is the (S) difference larger than the (R) and/or (RA) difference. Only 4 of 36 possible comparisons of (S) with (R) or (RA) do not favor the (S) condition reducing the achievement gap.

Table 7. Summary of Differences Between Free Lunch (FL) and Not Free Lunch (NFL) Students by Class Type (S, R, RA) on Percent of BSF Objectives Mastered by Years in a Class Type at Grades 1,2, and 3. (Percents Rounded).

| | | READING BSF (%) | | | MATH BSF (%) | | |
|---|---|---|---|---|---|---|---|
| | | S | R | RA | S | R | RA |
| Grade 1 | 1 yr (1) | 9 | 11 | 9 | 6 | 8 | 8 |
| | 2 yrs (K,1) | 11 | 15 | 10 | 8 | 11 | 9 |
| Grade 2 | 1 yr (2) | 7 | 14 | 15 | 3 | 7 | 8 |
| | 2 yrs (1, 2) | 11 | 14 | 12 | 4 | 7 | 7 |
| | 3 yrs (K, 1, 2) | 9 | 12 | 12 | 5 | 8 | 5 |
| Grade 3 | 1 yr (3) | 9 | 14 | 14 | 9* | 6 | 12 |
| | 2 yrs (2, 3) | 9 | 12 | 12 | 3 | 9 | 13 |
| | 3 yrs (1, 2, 3) | 16* | 13 | 13 | 20* | 15 | 10 |
| | 4 yrs (K, 1, 2, 3) | 11 | 9 | 9 | 8 | 9 | 10 |

* Only 5 of 36 comparisons of (S) with (R) or (RA) do not favor (S) in reducing the achievement gap. This grade 3, 3 years duration finding is consistent over all analyses, except the effect-size results found for norm-referenced tests (NRT) shown in Table 5.

Table 8. Percentage Retained in Grade by Class Type: Small (S), Regular (R) and Regular/Aide (RA), STAR Grades K-3. (Word et al., 1990, p. 170).

| | Percentage Retention | | | |
|---|---|---|---|---|
| Grade | Small (S) | Regular (R) | Reg./Aide (RA) | TOTAL |
| K | 3.8 | 4.5 | 3.7 | 4.0 |
| 1 | 7.8 | 12.6 | 10.8 | 10.6 |
| 2 | 4.7 | 5.6 | 4.0 | 4.7 |
| 3 | 3.5 | 6.7 | 4.0 | 4.1 |
| Ave over 4 grades | 4.9 | 6.8 | 5.7 | 5.8 |

Part 2 provides evidence of equity and achievement-gap aspects of small classes taken from long-term studies that have followed students into high school. The main body of the STAR cohort would have graduated in 1998 unless they moved or were retained in grade. Boyd Zaharias and Pate-Bain (2000) have been conducting STAR Follow-up studies to check on high school outcomes, drop out, completion of school with an honors diploma, and attending college in Tennessee.

Table 1 and Figures 1 and 2 on the following page (reproduced directly from original source and presented here collectively as Table 9) show the long-term results for some factors related to successful schooling. Effects of (S) will influence minority and non-minority students positively for a) graduating on schedule (the flip side of retention in grade), receiving an honors diploma, and dropping out of high school. Figure 1 shows the percent of STAR students attending college in Tennessee. The (S) advantage over (R) and (RA) for males, black students, black males, and (slightly) for free-lunch students is apparent. As with any cohort of 11,601 students, some will run afoul of the legal system. If this happened in Tennessee, Boyd-Zaharias and Pate-Bain obtained information about the STAR students' average prison sentence length, as shown in Figure 2. In all comparisons, students who had (S) classes in K-3 had shorter prison sentences. This was true for Black students, males, Black males, and students on free lunch.

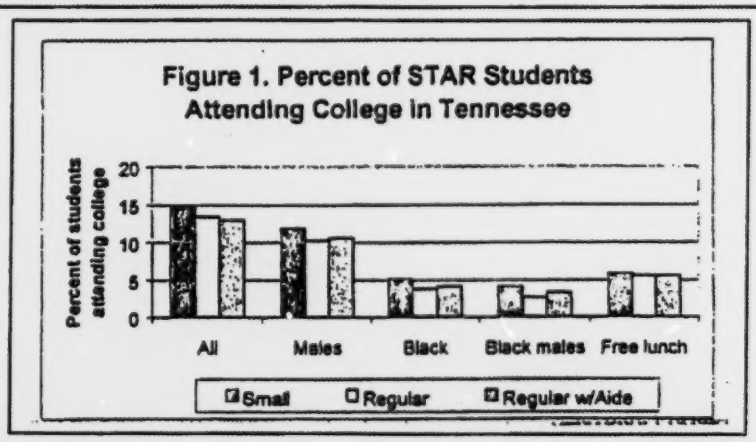
[TABLE 9 ABOUT HERE]

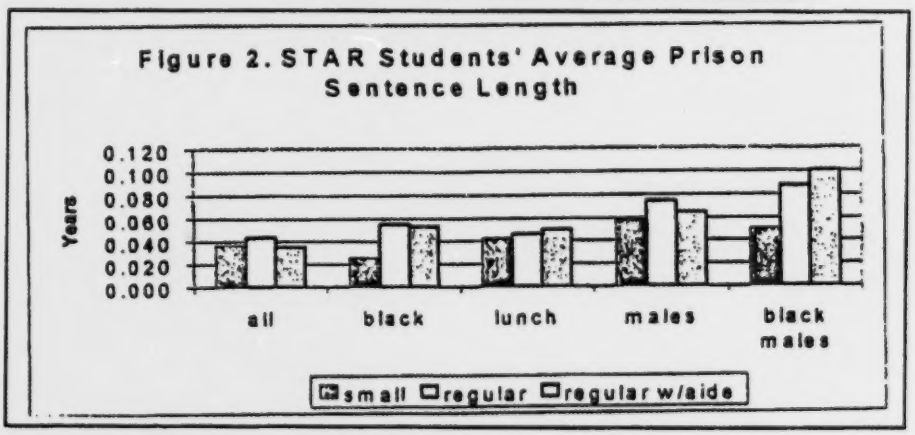
Table 9*.

Table 1. Effects of K-3 Small Classes on High School Graduation

| | Small | Regular | Regular/Aide | Total Sample |
|---|---|---|---|---|
| Available Records | 1652 | 1404 | 1821 | 4876 |
| Graduated on Schedule | **1256** (76.0%)* | **894** (63.7%) | **1274** (70.0%) | **3424** (70.2%) |
| Available Records | 1132 | 794 | 1124 | 3050 |
| Received an Honors Diploma | **510** (45.1%)* | **230** (28.9%) | **343** (30.5%) | **1084** (35.5%) |
| Available Records | 1506 | 1199 | 1630 | 4335 |
| High School Drop Outs | **228** (15.1%)* | **289** (24.1%) | **340** (20.9%) | **857** (19.8%) |

*Indicates the results were statistically significant ($p < .05$).




* Directly from Boyd-Zaharias and Pate-Bain (2000) pp. 10-11.

Some students who graduated (1998) took college-admissions test, either the SAT (a few) and the ACT (most of the students). The following material is taken verbatim as presented by Achilles and Finn (2000, March) at the American Educational Research Association (AERA), 2000 (p. 5). The original source is Krueger and Whitmore (1999).

> Krueger and Whitmore (1999) considered the effect of attending a small class in the early grades (K-3) on college attendance plans. They reported results based upon:
>
> > . . . the percent of students who took either the ACT or the SAT exam by the type of class they were assigned to attend their initial year in Project STAR. The figures are reported for all students combined, for white and black students separately, and for students who received free or reduced-price lunch in at least one year in grades k-3. For the entire sample, . . . 43.7% of students who were assigned to a small class took either the ACT or SAT exam, whereas 40.0% of those assigned to a regular-size class took one of the exams, and 39.9% of those assigned to a regular-size class with an aide took one of the exams. The 3.6 percentage point higher test-taking rate for the students assigned to small classes relative to those assigned to regular-size classes was statistically significant at the .05 level; that is, this difference is unlikely to have occurred by chance.
> >
> > 6. . . attending a small class was particularly effective in raising the proportion of black students who wrote one of the college entrance exams. Only 31.7% of black students in regular-size classes wrote the ACT or SAT exam, whereas 40.2% of black students in small classes wrote the college entrance exam. To gain some perspective on the magnitude of this effect, note that black-white gap in taking a college entrance exam was 13.3 percentage points for students in regular-size classes, and 6.1 percentage points for students in small classes. Thus, attending a small class reduced the black-white gap in the college-entrance-test-taking rate by 54 percent.
> >
> > 6 Earlier research on Project STAR has found that minority students and students on free lunch exhibited the greatest gains in test scores as a consequence of attending a small class. The findings here complement a result that has been found consistently throughout Project STAR: minority students benefited most from attending a small class, and small classes were able to considerably narrow, although not eliminate, the gap in educational performance between black and white students. (These results support Robinson's [1990] findings from his review of the early class-size research),

Data from Tennessee's Project STAR clearly show that small classes in elementary grades (K-3) help to reduce the traditional "gaps" between minority and non-minority youth in schools. In grades K-3 there are test-score data on the achievement gap from NRT, but mostly from CRT. There also are data on retention in grade. Data from later years of the students' lives emphasize more the social benefits and high school completion rates, but the achievement gap is always reduced for students who had been in (S) when compared to students who had been in (R) or (RA) in grades K-3.

## Section III. Evidence of Gap Reduction from Non-STAR Sources

This section includes some non-STAR sites that have indications that a reasonable class size may have been a factor not only in absolute gains in test-score results, but also in "achievement-gap" change. At best, the conclusions here are based on observed relationships as interpreted by researchers on class size in the early grades of schooling. The STAR results serve as "benchmarks" for comparisons. For example, in STAR the minority students (mostly African American) received two or three times the benefit from small classes that non-minority students received. By way of comparison, in the Texas example the change or gain for white students of 25% and for economically disadvantaged students of 55% from 1994 to 199 would seem "within the ball park" based on STAR results. (Not Tabled).

The Fairfax County Public Schools (FCPS) Virginia Class-Size Evaluation.

The Fairfax County (VA) Public Schools (FCPS) began a small-class effort with planing in 1991 and implementation in 15 schools in Fall, 1992. By the 1996-97 school year 48 schools were participating. Although the Final Report (FCPS, 1997) does not show students by race, it does report data related to student socio-economic status (SES), and the achievement gap issue is relevant to SES, as well as to gender, race, and ethnicity. The class-size effort began in grade 1, in schools with the highest percentage of low-SES pupils. The Executive Summary (FCPS, 1997) highlights findings as follows.

> The supporting data indicated strong evidence that the Reduced-Ratio Program is more effective for students from low (SES) backgrounds. There is also evidence that ideal program implementation is an important factor in increasing student achievement ... (ES-5, Emphasis Added).
>
> Consider extending the program beyond first grade in the most needy schools . . . FCPS students from low (SES) backgrounds and students in schools with ideal implementation showed some significant gains on a nationally norm-referenced test three years after participating in the program. (ES-5, Emphasis Added).

An important finding from the FCPS study is the difference in results that is apparent in

the comparison of test-score results (1993-94 to 1995-96) on Grade 4 ITBS between the 10 schools with "ideal" implementation and the 19 schools that did not have ideal implementation. In the ideal implementation, there were positive ITBS gains on 6 of 8 subtests (the other two had positive gains), and in the other 19 schools there were no gains on 6 of 8, a positive gain on one, and a loss on one. Table ES-3 from the FCPS study is reproduced below (p. ES-4) as Table 10.

---

Table 10.

**Table ES-3**
**Average Scale Score Change on 1995-1996 Grade 4 ITBS**
**When Compared to 1993-1994 Scores**

| ITBS Subtest | 10 Schools with Ideal Implementation | 19* Schools with Variations in Implementation |
|---|---|---|
| Reading Comprehension | +3** | 0 |
| Vocabulary | +2 | -1 |
| Total Language | +4** | 0 |
| Math Concepts | +5** | +1 |
| Math Problem Solving | +2 | 0 |
| Math Computation | +2** | 0 |
| Total Math | +3** | 0 |
| Composite | +3** | 0 |

*Markham and Cheney were not included in this analysis because there are no fourth grades in these schools.
**Significant $p<.05$

This FCPS information supports the contention that there are right (ideal) ways to implement class-size changes in a school or school system in order to get optimum student gains. Detailed and research-based guidelines on how to implement small classes have been described elsewhere [e.g., Achilles & Finn (2000, March)]. Key points are: 1) Early Intervention, 2) Intense—all day, each day, and 3) Sufficient Duration:

1. Start the small class when the pupil starts school (K) or (Pre-K).
2. Avoid PTR-like events (e.g., pull outs). Keep the class together with the teacher.
3. Maintain the small class (15-18 to 1) for at least 3, preferably 4 years.

The Burke County, North Carolina, Class-Size Effort.

Burke County is a relatively low-wealth system located in Western NC in the foothills of the Great Smoky Mountains. As it has grown, the school system has experienced an influx of limited English proficient (LEP), English as a Second Language (ESL), and low SES youngsters. Table 11 shows the changes, 1992-1993 through 1997-1998: school population increase of 11% (12,400-13765), Free lunch increase of 45%, LEP increase of 423%. Title-I schools where at least 50% of the students qualify for Title I increased from one (1995-95) to five (2000); and seven other schools are served with targeted Title I efforts because the percentage of eligible youngsters is equal to or greater than the county average.

To counter low student test scores, Burke County Schools (BCS) initiated class-size reductions in grade one of four pilot schools (1991-92). Results were good, so in 1992-93 small classes (about 15:1) were begun in grade one of all 14 elementary schools and in grade two at the four pilot schools. By 1994-95 the initiative was in grades 1 and 2 at all 14 elementary schools, and by 1995-96 into grades 1-3 at all schools. This small-class phase-in was done with local funds by creative reallocations and re-assignments in a low-wealth district. (Stewart, 1998: Achilles, Egelson, & Harmon 1995).

Table 11. Student Demographic Changes, Burke County, NC 1992-93 to 1998-98: Population, Limited English Proficiency (LEP), Free Lunch (FL) and Title I Eligible.

| | 1992-93 n | 1997-98 n | Growth n | Growth % |
|---|---|---|---|---|
| (A) Population | 12,400 | 13,765 | 1365 | 11 |
| (B) Free Lunch | 3583 | 5199 | 1616 | 45 |
| %B/A | 29 | 38 | | |
| (C) Limited English Proficient (LEP) | 217 | 1135 | 918 | 423 |
| % C/A | 2 | 8 | | |

Most students who started school in BCS in grade one in 1992 and all students who started in 1993-94 would have been in small classes, grades 1, 2, and 3. Measures reported here are results from North Carolina's end-of-grade (EOG) tests. The percent of grade-level students "proficient" by reason of scoring in levels 3 or 4, the top two levels of the EOG proficiency tests. Level 1, the bottom, includes students who are the least proficient, and who most need growth. A learning-achievement goal, then, is to greatly reduce or diminish the percentage of students scoring in level 1 and to increase the percentage of students scoring proficient, or in levels 3 and 4. By 1994-95 about half (most) of grade-3 students had small-classes in grades 1-3. By 1995-96 all BCS students had small classes in grades 1-3 when they took the Grade-3 EOG tests. Table 12 shows both the decline in level 1 and the increase in levels 3 and 4 (proficient) in the EOG tests for 1994-95 and 1995-96, a direct result of the small-class effort. The 1998 through 2000 results show that this increase was maintained despite the 11% increase in population, 45% increase in free lunch, 423% increase in LEP, and a large increase in Title-I students served.

Table 12. Burke County, NC EOG TEST Outcomes. Grade 3 (Stewart, 1998 & updated)*

| YEAR that (MOST) or ALL students taking EOG Test could have had grades 1-3 in small classes. | Level 1 % Not Proficient | | Levels 3 & 4 % Proficient | |
|---|---|---|---|---|
| | Reading | Math | Reading | Math |
| 1993 | 13.4 | 10.9 | 60.3 | 61.6 |
| 1995 (MOST) | 8.6 | 5.6 | 68.6 | 71.4 |
| 1996 ALL | 5.9 | 4.7 | 74.0 | 75.9 |
| 1998 ALL | 4.4 | 2.4 | 79.8 | 76.0 |
| 1999 ALL | 4.0 | 2.9 | 78.1 | 75.7 |
| 2000 ALL | 2.4 | 2.1 | 83.9 | 80.5 |

Table 13 shows the long-term impact of the early (grades 1-3) small-class start in school as shown by the EOG tests, percent proficient in grades 5-6-7-8 for 1992-93 through 1999-2000. The grades 1-3 small-class implementation was completed in 1995-1996 and the student cohort who had small classes in grades 1-3 moved through the tests at grades 5, 6, 7, and 8 as shown here in two groups: about half or most students in small classes (most), and all students in small classes (all):

1996-97 [Grade 5 (most); 6-8 (none)]
1997-98 [Grade 5 (all); 6 (most); 7-8 (none)]
1998-99 [Grades 5-6 (all); 7 (most); 8 (none)]
1999-00 [Grades 5-7 (all); 8 (most)]

* Thanks to Burke County, NC personnel for compiling and sharing these results, especially former superintendent T. Stewart, and present personnel M. Church, S. Schwengel, S. Wilson, and T. Farris.

Table 13.

**NORTH CAROLINA END-OF-GRADE TESTING PROGRAM**
**JUNE 2000**

**SCHOOL: BURKE COUNTY - MIDDLE** PRELIMINARY

**MATH – PERCENT OF STUDENTS PROFICIENT**

| | 1992-93 | 1993-94 | 1994-95 | 1995-96 | 1996-97 | 1997-98 | 1998-99 | 1999-2000 |
|---|---|---|---|---|---|---|---|---|
| GRADE 5 | 62 | 67 | 72 | 78 | 77 | 85 | 87 | 89 |
| GRADE 6 | 66 | 67 | 72 | 76 | 78 | 86 | 85 | 88 |
| GRADE 7 | 59 | 66 | 67 | 67 | 75 | 83 | 88 | 85 |
| GRADE 8 | 67 | 64 | 66 | 65 | 66 | 81 | 82 | 85 |

**READING – PERCENT OF STUDENTS PROFICIENT**

| | 1992-93 | 1993-94 | 1994-95 | 1995-96 | 1996-97 | 1997-98 | 1998-99 | 1999-2000 |
|---|---|---|---|---|---|---|---|---|
| GRADE 5 | 70 | 70 | 73 | 73 | 74 | 83 | 79 | 82 |
| GRADE 6 | 66 | 66 | 67 | 72 | 73 | 74 | 78 | 75 |
| GRADE 7 | 66 | 69 | 70 | 67 | 73 | 77 | 81 | 79 |
| GRADE 8 | 71 | 74 | 78 | 71 | 75 | 83 | 85 | 89 |

---

The percentages of students proficient in math and in reading increase generally as the cohorts with about half (most) and (all) students move through grade–level testing sequences. The increase in percent proficient that occurs in 1997-98 will include the first cohort of small-class students who began in the pilot schools in 1991-92 (grade 7), the cohort of small-class students most of whom had 1-3 small classes (grade 6) and the cohort (grade 5) who all had small classes in grades 1-3. The grade 8 increase (1997-98) needs more analyses: it does not seem related to small classes in grades 1-3.

Indicators of Gap Reduction from Texas Data

In 1984 Texas passed HB 72 which mandated small classes, not to exceed 22 students in grades K-2 in all Texas schools by 1985-86, and in grades K-4 by 1988-89 (Part of HB 72 is in Appendix C). Thus, a Texas student who began school in 1985 could have experienced small classes in grades K-4. Student academic progress in Texas is measured several ways. One statewide test is the Texas Assessment of Academic Skills (TAAS). The first year that grade-10 students who could have had small classes in grades K-2 took the TAAS was 1995-96. Grade-10 students who took the TAAS in 1995-96 could have remained in small classes for five years (K-4), but if there was any lag time in full implementation of small classes, the small-class effect would show up in later years of testing (1997-98, 1998-99, 1999-00) and should level off unless there is an increase in pre-K interventions, or for other reasons described in Appendix D.*

Figure 1 is an adaptation of data presented by Cortez (1999, pp. 5-6). The percentage of All Texas students passing all three parts of the TAAS increased from about 50% before HB 72 to 75% by 1999, a 25 point gain. The percentage of economically disadvantaged students passing the TAAS rose from 30-34% before HB 72 to 68% by 2000, a 34-38 point gain. The gap reduction begins with the 1995-96 testing, the first year that sophomores (grade 10) would have experienced HB 72 small classes that began in 1985-86.

Figure 2 (adapted from Cortez, 1999) shows scores of three groups (White, African-American, and Hispanic) on TAAS, 1991-92 – 1999-00. Before small classes, the pass rates were in the 60's for White and in the 28-36 range for minorities, about a 30-point gap. By 2000 the score for Whites was 89 and the minority scores were 67-70, a 20-point gap. The slope of the lines and the gap reduction start with the 1995-1996 testing, when sophomores would have had small classes.

---

* Appendix D provides analyses by Haney (2000) that identify large increases in grade retention, dropouts, and special education exemptions, especially for minority students.

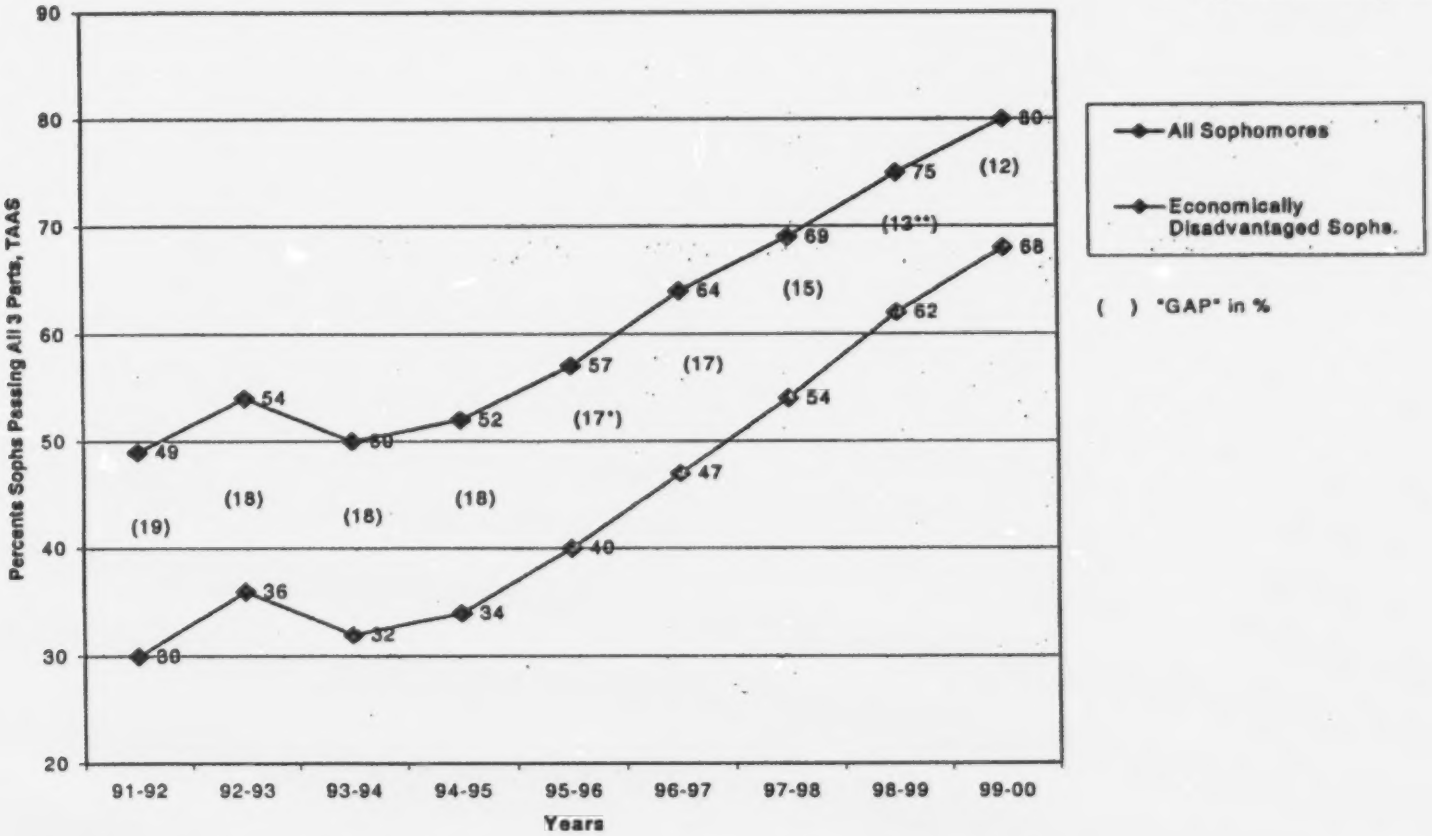
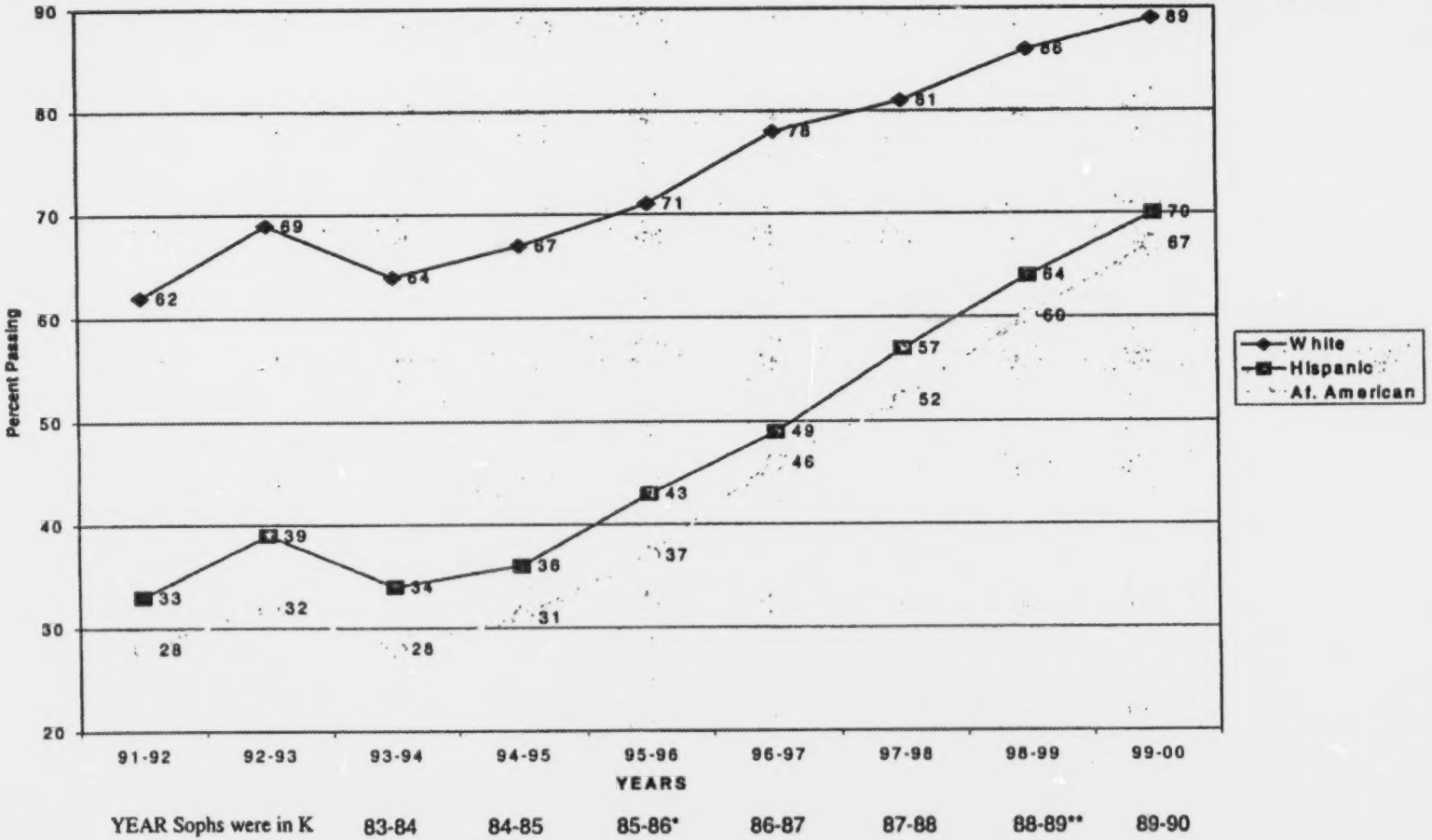



Figure 1. Percent (%) Sophomores Passing All 3 Parts of TAAS Related to Implementation of House Bill (HB) 72 in Texas that Required Class Sizes, K-2 (1985-86) and then K-4 (1988-89) Not to Exceed n=22.




* Implement HB 72 in K-2 ** Implement HB 72 in K-4

Figure 2. Percent (%) Sophomores Passing All 3 Parts of TAAS by Three Race/Ethnic Groups [Adapted from Cortez (1999)].

Table 14 compares the percent of grade-10 students passing the TAAS by selected groups of students: White, Hispanic, African-American, Economically disadvantaged, and All students, by years (1993-94 through 1998-99). Table 14 data seems to show the impact of small classes on the differing groups with minority and economically disadvantaged students getting larger gains than do white students, and that the largest gains occur after the sophomore test takers could have had at least two years in a small class (1996-97).

Table 14. Percent of Groups of Grade-10 Students Passing All Three Parts of TAAS by Years. (Adapted from Cortez, 1999, pp. 6-7).

| | Percent Passing by Groups. ( ) = Difference from White | | | | | |
|---|---|---|---|---|---|---|
| YEAR | ALL | WHITE | HISP* | AF. AM. * | ECO. DIS. * | Possible Years (n) in Small Class (K-4) ** |
| 93-94 | 50 | 64 | 34 (30) | 28 (36) | 32 (32) | 0 |
| 94-95 | 52 | 67 | 36 (31) | 31 (36) | 34 (33) | 0 |
| 95-96 | 57 | 71 | 43 (28) | 37 (34) | 40 (31) | 1 |
| 96-97 | 64 | 78 | 49 (29) | 46 (32) | 47 (31) | 2 |
| 97-98 | 69 | 81 | 57 (24) | 52 (29) | 54 (27) | 3 |
| 98-99 | 75 | 86 | 64 (22) | 60 (26) | 62 (24) | 4 |
| 99-00 | 80 | 89 | 70 (19) | 67 (22) | 68 (21) | 5 |
| GAIN | 30 | 25 | 36 (11) | 39 (14) | 36 (11) | |
| Diff from ALL | | -5 | +6 | +9 | +6 | |

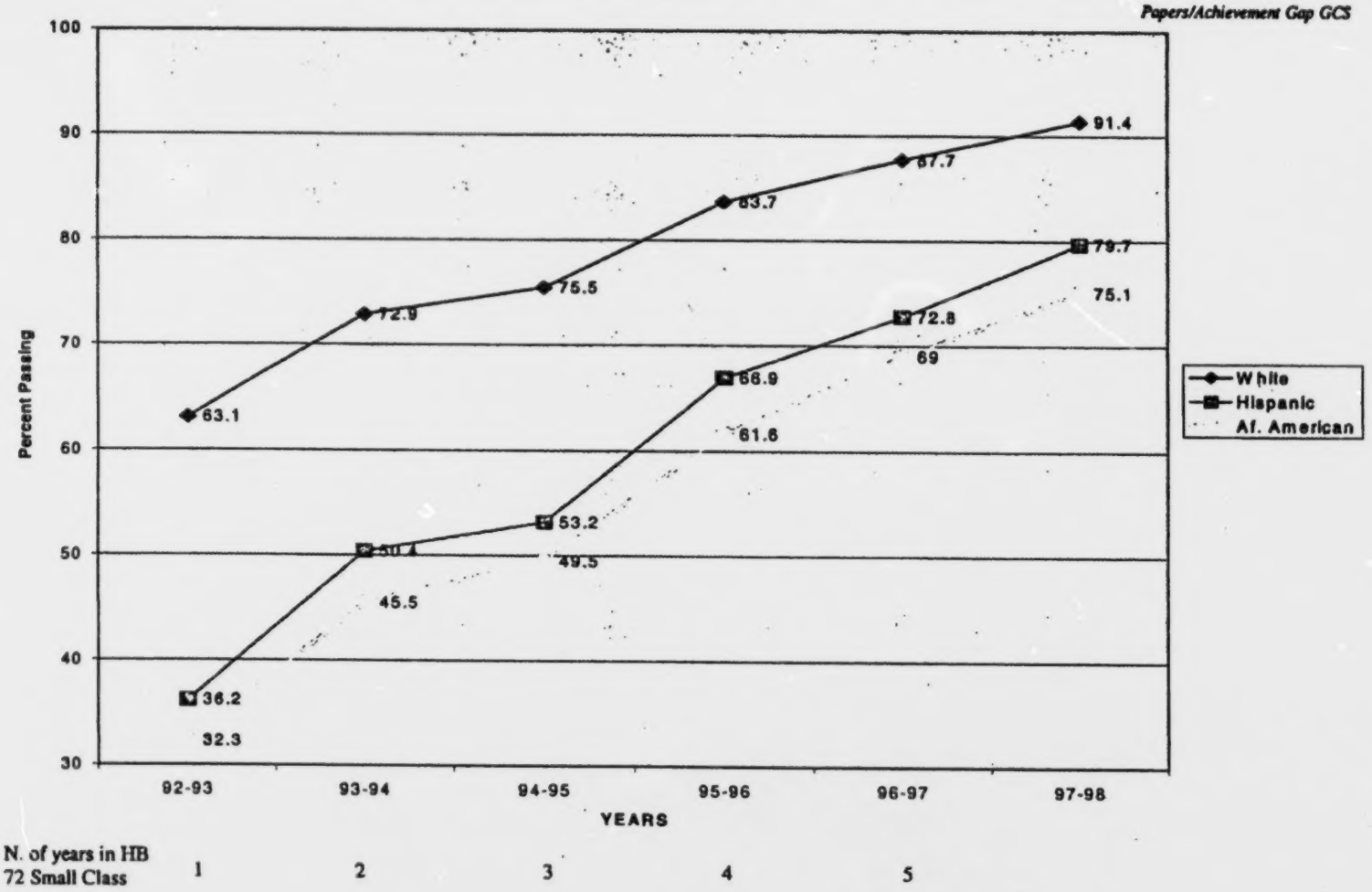
* Hisp = Hispanic; AF. AM. = African American; ECO. DIS. = Economically Disadvantaged

** NOTE. In 1993-94 grade-10 students would have had no years of HB 72 small classes. Project STAR analyses showed that 3 years or more in a small class would have lasting benefits into grades 8, 10 and beyond.

Figure 3 provides data from Navarro and Natalicio (1999) showing the TAAS passing percentages for three groups of students in El Paso, TX from 1992-93 through 1997-98. Of particular interest is the evident gap reduction between the white students and the Hispanic and African-American student groups, as well as the commendable increase of all groups in the percent passing. White student passing rates went from 63.1 to 91.4 (28.3 percentage point gain), but Hispanic rates moved from 36.2 to 79.7 (43.5 gain) and African-American rates increased from 32.3 to 75.1 (42.8 gain). The "gap" closed for Hispanics from 26.9 to 11.7 percentage points, and for African American students from 30.8 to 16.3. The largest gains occurred after 1994-95 when students taking the test could have had HB 72 small classes.

Because of Texas' early move to small classes, K-4, the state's data seem to provide evidence over time both of the small-class effect and how small classes reduce the achievement gap over time. The magnitude of the state-wide gains, and the similar gains in one area (El Paso) reflect the gains and the gap-reduction evidence obtained from Project STAR. Grissmer et al., (2000) praised Texas as one state that has made progress on the National Assessment of Educational Progress (NAEP), and, according to the RAND Study, states that made progress have, among other things, small classes in elementary schools and public pre-kindergarten programs, two components of the Texas education reform started with HB 72 in 1984.*

* Appendix D provides some indications that forces such as grade retention, dropouts, and exemptions from test taking by qualifying for special education need to be thoroughly considered to see if the "Texas Miracle" and be a "Mirage." (Haney, 2000).




Implement HB 72 in K-2 (1985-89) and in K-4 (1988-89).

Figure 3. Percent (%) Sophomores Passing TAAS Math Exams. (Adapted from Navarro and Natalicio, 1999. Kappan).

## Section IV. Summary, Conclusion, Conjectures, Recommendations

This paper is a "first pass" at exploring evidence that small (S) classes in early elementary grades (K-3 or K-4) will help to reduce the achievement gap (variously defined) between and among selected groups of students. Data were drawn from Project STAR and its related studies, and from selected other class-size initiatives and evaluations. One finding, that the "small-class effect" is not distributed equally to all students but that it has a stronger impact on non-majority students, is consistent here with studies by Robinson (1990), Wenglinsky (1997), SAGE evaluators (Molnar et al., 2000), and Achilles, Finn, and Bain (1997/1998) on the same topic. Small classes employed equally provide an equity factor.

Based upon the consistency of all of these findings, a major conclusion is that all youngsters deserve the proven benefits of starting their schooling in small classes of approximately 14-18 youngsters and one qualified teacher. During these early years (K or pre-K to grade 3 or 4) the students need to be assigned to classes at random (heterogeneously) as in STAR to benefit the most from peer interactions. They should generally move in a cohort (as in STAR) to benefit from a sense of community. If the group stays with the same teacher for two or more years, this would be looping, and presumably this arrangement, along with elements of non-gradedness, will add to the small-class and K (or pre-K) experiences.

Various media outlets emphasize the fairly wide-spread concerns about the need for education practitioners and researchers to continue to study and diminish the "achievement gap." Yet, it is reasonable to expect that at some level, there will continue to be achievement gaps if we believe in individual differences. If educators teach all children to read and to work with numbers, assuredly some children will read or cipher more and better than others. There will continue to be group differences that can be in impelled by major detriments to learning, such as

poverty, hunger, opportunities for early intervention, etc. As with the "Effective Schools" ideas, we must believe, however, that we can and should dramatically reduce differences between groups in disaggregated data of basic skills and knowledge taught during schooling.

The research on class size, and especially STAR and its related studies, has provided key information on how small classes should be implemented for minimum confusion during implementation and for maximum gains during operations: Start early, when the child enters K (or pre-K) and add one grade per year. (STAR analyses consistently show that one year of a small class in grade 3 did not provide short-term or long-term results. For lasting or enduring benefits, students must remain in (S) for at least 3, and preferably 4 or 5 years in a true class-size reduction, not in a PTR melange of pull-outs, teacher aides, and other activities that take a student from the teacher and peers and that reduce time on task (while the student wanders the halls).

Information about HOW to implement small classes in schools and WHY small classes work is becoming more and more available as more schools and systems implement small classes and research and evaluation studies are conducted. Some important findings about implementing small classes are shown here.

A. Small Classes Need to be Systematically Initiated and Used According to the Research Findings

1. Early Intervention. Start when the pupil enters "schooling" in K or even pre-K.
2. Intense Treatment. All day, every day in the small class. Avoid Pupil-Teacher Ratio (PTR) events, such as "pull-out" or team teaching.
3. Sufficient Duration. Maintain the small class for at least 3, preferably 4, years for enduring effects.
4. Use random assignment in early grades (STAR) to facilitate peer tutoring and problem-solving groups.
5. Employ a cohort model (STAR) for several years so students (and teachers develop a sense of family or community.

B. Positive Benefits of Small Classes Are Supported by Many Well Researched Theories About Teaching, Learning, and Learning Groups.

1. Time On Task
2. Group Dynamics
3. Engagement/Participation
4. Peer Interaction/Tutoring
5. Task Induction
6. Immediate Reinforcement
7. Classroom Environment (e.g., Air Quality, Materials, Space/Crowding)
8. Classroom Management
9. Increased Parent Interest
10. Individualization
11. More Teaching Methods
12. Teach To Mastery
13. Personal Attention/Family
14. Teacher/Student Morale
15. Early Diagnosis of Learning Difficulties
16. Less Indiscipline
17. Responsibility/Accountability

C. Small Classes Facilitate Important Class-Level Activities

1. Inclusion Efforts
2. Field Trips, Celebrations and Other Events
3. Parent Night/Parent Involvement
4. Student-Led Activities

Results shown here, from STAR and from other studies, seem to extend work showing the importance and benefits of small-class early intervention achieved by such longitudinal and long-standing work as Head Start, Abecedarian (Ramey & Ramey, 1998) and the venerable Perry Pre-School Project. Convergence of these results adds a sense of validity and common-sense agreement to all of the research.

We'll continue to compile evidence to see if and how small classes in elementary grades help reduce the academic achievement gap. As this work proceeds, ideas, issues, and questions

continue to arise. For example, in spite of the achievement-gap reductions that are related to small classes, many other questions deserve research attention and should be studied.

A conjecture. Research has shown the positive impact of a mother's education on the education outcomes of children. As the general education level of the U. S. population rises, and as more and more mothers attend post-secondary school, will this influence an academic achievement-gap decline. Put into question form: How it is achievement-gap reduction related to the increasing levels of parent education, especially of the mother?

## REFERENCES

Abt Associates (1997, April). Prospects: Final Report on Student Outcomes. Cambridge, MA: Author. Report prepared for U.S. Department of Education.

Achilles, C. M. (1999). Let's Put Kids First finally: Getting Class Size Right. Thousand Oaks, CA: Corwin Press.

Achilles, C. M. (1998, April). Small for all. Phi Delta Kappan, 79 (8), 640.

Achilles, C. M., & Finn, J. D. (2000). Should class size be a cornerstone for educational policy? CEIC Review, Philadelphia, PA: Temple University, Center for Education in the Inner Cities. 9 (2), 15, 23.

Achilles, C. M., Finn, J. D., & Bain, H. (1997-1998). Using class size to reduce the equity gap. Educational Leadership, 55 (4), 40-43.

Achilles, C. M., Harman, P. & Egelson, P. (1995, Fall). Using research results on class size to improve pupil achievement outcomes. Research in the Schools, 2, (2), 23-30. (Paper by same title presented at AASA convention, 2/95).

Achilles, C. M., Kiser-Kling, K., Owen, J., & Aust, A. (1994). Success Starts Small: Life in a Small Class. Greensboro, NC: University of North Carolina at Greensboro, Final Report. Small-grant/School-based Research Project. (ERIC EA 029049)

Achilles, C. M., Nye, B. A. & Bain, H. P. (1994-95). The test-score "value" of kindergarten for pupils in three class conditions at grades 1, 2, and 3. National Forum of Educational Administration and Supervision Journal, 12(1), 3-15.

Akerhielm, K., (1995). Does class size matter? Economics of Education Review, 14 (3), 229-241.

Archibold, R. C. (2000, September 3). Much talk of education on the stump, much skepticism in the schools. The New York Times: New York Report, p. V27.

Boozer, M. & Rouse, C. (1995, May). Intraschool variation in class size: Patterns and implications. (ED 385935). Princeton, NJ: Princeton University, Industrial Relations Section Paper #344. National Bureau of Economic Research, Cambridge, MA.

Borman, G. D. & D'Agostino, J. V. (1996, Winter). Title I and student achievement: A meta-analysis of federal evaluation results. Educational Evaluation and Policy Analysis, 18 (4). 309-326.

Boyd-Zaharias, J. and Pate-Bain, H. (2000, April). The continuing impact of elementary small classes. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA. Lebanon, TN: HEROS, Inc.

Butzin, S. (2000, September). Iceberg ahead. Backtalk. Phi Delta Kappan, 82 (1), 98.

Cortez, A. (2000, March). Why better isn't enough: A closer look at TAAS gains. IDRA Newsletter. San Antonio, TX.

Fairfax County (VA) Schools (1997, July). Evaluation of the reduced-ratio program: Final report. Fairfax, VA: Office of Program Evaluation, Fairfax County Schools. (J. DiStefano, Principal Investigator).

Finn, J. D. (1998, April). Class size and students at risk: What is known? What is next? Washington, DC: U.S. Department of Education. Office of Educational Research and Improvement (OERI), AR 98 7104.

Finn, J. D., & Achilles, C. M. (1999, Summer). Tennessee's class size study: Findings, implication, misconceptions. Educational Evaluation and Policy Analysis, 21 (2), 97-107.

Finn, J. D., & Achilles, C. M. (1990, Fall). Answers and questions about class size: A statewide experiment. American Educational Research Journal, 27(3), 557-577.

Finn, J. D., Gerber, S., Achilles, C. M., & Boyd-Zaharias, J. (1999, May). Short and long-term effects of small classes. Paper presented at American Educational Research Association Annual Meeting, New Orleans, LA.

Finn, J. D., Gerber, S., Farber, S. L. & Achilles, C. M. (2000, March). Teacher aides: An alternative to small classes? CEIC Review, Philadelphia, PA: Temple University, Center for Education in the Inner Cities. 9 (2), 5.

Grissmer, D., Flanagan, A., Kawata, J. & Williamson, S. (2000). Improving Student Achievement: What State NAEP Test Scores Tell Us. Santa Monica, CA: RAND Education.

Krueger, A. B. (2000, March). An economist's view of class size research. CEIC Review, Philadelphia, PA: Temple University, Center for Education in the Inner Cities. 9 (2), 19-20.

Krueger, A. B. (1998, January). Reassessing the view that American schools are broken. Princeton, NJ: Princeton University, Industrial Relations section working paper 395. (1998, March). FRBNY Economic Policy Review, 29-43.

Krueger, A. B. (1997, June). Experimental estimates of education production functions. Princeton, NJ: Princeton University, National Bureau of Economic Research Paper #6051. Cambridge, MA.

Krueger, A. B. & Whitmore, D. M. (2000, March). The effect of attending a small class in the early grades on college-test taking and middle school test results: Evidence from Project STAR. Princeton University.

Krueger, A. B. & Whitmore, D. M. (1999, April). The effect of attending a small class in the early grades on college attendance plans. Executive Summary. Princeton University and NBER.

Levin, H. M. (1998, May). Educational performance standards and the economy. Educational Researcher, 27, (4), 4-10.

Molnar, A., Smith, P., Zahorik, J., Palmer, A., Halbach, A., & Ehrle, K. (2000, March). Wisconsin's Student Achievement Guarantee in Education (SAGE) class-size reduction program: Achievement effects, teaching and classroom implications. CEIC Review, Philadelphia, PA: Temple University, Center for Education in the Inner Cities. 9 (2), 12-13.

Molnar, A., Smith, P., Zahorik, J., Palmer, A., Halbach, A., & Ehrle, K. (1999, Summer). Evaluating the SAGE program: A pilot program in targeted pupil-teacher reduction in Wisconsin. Educational Evaluation and Policy Analysis, 21 (2), 165-178.

Mosteller, F. (1995). The Tennessee study of class size in the early school grades. The Future of Children, 5 (2), 113-127.

Mosteller, F., Light, R. J., & Sachs, J. A. (1986, Winter). Sustained inquiry in education: Lessons from skill grouping and class size. Harvard Educational Review, 66 (4), 797-828.

Navarro, S. & Natalicio, D. S. (1999, April). Closing the achievement gap in El Paso. Phi Delta Kappan, 80 (8), 597-601.

Pogrow, S. (2000, September). Success for All does not produce success for students. Phi Delta Kappan, 82 (1), 67-80.

Pogrow, S. (2000, April). The unsubstantiated "success" of Success for All. Phi Delta Kappan, 81 (8), 596-600.

Ramey, C. T., & Ramey, S. (1998). Early intervention and early experience. American Psychologist, 53, 109-120.

Ritter, G. W. & Boruch, R. F., (1999, Summer). The political and institutional origins of a randomized controlled trial on elementary school class size: Tennessee's project STAR. Educational Evaluation and Policy Analysis, 21 (2), 111-126.

Slavin, R. E. (2000, May). Research overwhelmingly supports Success for All. Backtalk. Phi Delta Kappan, 81 (9), 720.

Slavin, R. E. & Madden, N. A. (2000, September). Research on achievement outcomes of Success for All. Phi Delta Kappan, 82 (1), 38-40, 59-66.

State of Texas (1984). General and special laws of the State of Texas. Sixty-eighth Legislature 6/4/84-7/11/84. Esp. Chapter 28, H.B. No. 72.

Stewart, T. (1998, June 29). Reduced size classes. Data Tables and Narrative. Morganton, NC: Mimeo. Updated (2000) with information provided by M. Church, Testing and Evaluation, Burke County, NC.

Voelkl, K. (1995). Identification with school. Unpublished PhD dissertation. Buffalo, NY: SUNY. (UMI #0538143).

Wong, K. K. & Meyer, S. J. (1998, Summer). Title I schoolwide programs: A synthesis of findings from recent evaluations. Education Evaluation and Policy Analysis, 20 (2), 115-136.

Word, E., Johnston, J., Bain, H., Fulton, B., Zaharias, J., Lintz, N., Achilles, C. M., Folger, J., & Breda, C. (1990). Student/teacher achievement ratio (STAR): Tennessee's K-3 class size study. Final report and final report summary. Nashville, TN: Tennessee State Department of Education.

# Appendix A

## A Longitudinal Class-Size Experiment.

Project STAR (1985-1989) and the many studies that build upon STAR (See Table 1 for a partial listing) benefit from the experiment's tightly controlled, in-school longitudinal design. STAR was conducted by a four-university consortium with considerable external support from consultants, advisory groups, and the Tennessee State Department of Education. Basic design issues are:

(1) Project STAR built on principles recognized in prior research. The intervention began in the primary grades. Small classes had fewer than 20 students. STAR's design enabled researchers to look at the effects on minority as well as majority students. Moreover, the design required a "real" difference in the class sizes, from an average of 24 pupils to an average of 15.

(2) STAR was a controlled experiment that permitted, to the extent possible with empirical data, causal conclusions about outcomes. Pupils entering K were randomly assigned to a small class (S; 13-17), a regular class (R; 22-27), or a regular class with a full-time teacher aide (RA). Pupils entering in later years were also assigned at random to classes. Teachers were assigned to classrooms at random. Randomization was monitored carefully.

(3) With minor exceptions, students were kept in their class grouping in grades K, 1, 2, and 3. A new grade-appropriate teacher was assigned to the class each year. STAR was a four-year longitudinal experiment.

(4) Norm-referenced tests (NRT), and criterion-referenced tests (CRT) and measures of self concept and motivation were administered each spring. Researchers used a post-test only design. (Campbell and Stanley, 1963).

(5) The samples were large and diverse. The K year involved over 6300 students in 329 classrooms in 79 schools in 46 districts. The first-grade sample was larger still. The large samples were maintained throughout the four years, producing an excellent longitudinal database. Total sample = 11,601.

(6) The class arrangement was maintained throughout the day, all year long. There was no intervention other than class size and teacher aides. No special training was provided to the teachers except for a small sample in second grade; no special curricula or materials were introduced. (Training didn't increase outcomes).

(7) Students were followed and evaluated after STAR ended in grade 3. Most students graduated in 1998. Their college-entrance test results were monitored. (Krueger & Whitmore, 1999).

## APPENDIX B

A Summary of Class-size Findings, and Issues Derived from Various Sources.*

| Finding, Idea, Issue, or Question | | Selected Sources of Support |
|---|---|---|
| I. | Class-size effect was found in all sites, for all participants, at all times and grades K-3, This includes tutoring and "special" projects. | STAR, Challenge, Reading Recovery (RR); Success for All (SFA) |
| II. | Small classes work best when student starts (K, 1) school in them; they are preventive, not remedial. Compelling research suggests that formal and small-class education MUST start no later than K. | STAR, SSS, Challenge SAGE, Burke Co. Abecedarian (NC), Finn & Achilles (1999) |
| III. | Crowding, not just small classes, is an issue. School safety and environment are improved. (Prout, In process). School size is related. | STAR, SSS, Nye, Fowler & Walberg, Behavioral Research, Cotton, others. |
| IV. | Although all pupils benefit from small (S) classes in K-3, some students benefit more than others. | STAR, SFA, RR, LBS, Other class-size work |
| V. | The teacher is important. Each pupil's learning depends upon the teacher and others in the class. (Thus the class is the unit of analysis). | STAR, LBS, SSS, Challenge, Burke County, CSR in California. |
| VI. | A teacher aide does not improve student outcomes. This adds to crowdedness and causes new dynamics (Issues: Training, inclusion, ESL, role description). | STAR, Other Studies of Teacher Aides, Finn & Gerber (In process) |
| VII. | Teachers should use known educational-improvement process: (Parent and home involvement, portfolios, alternative assessments, etc.). Small classes do not change what teachers do—just how much they do good things well. | STAR, LBS, SSS, Filby et al., Burke County, NC Downtown School, NC STAR Teacher Studies. |
| VIII. | Reduce retention in grade especially when student will be moving into another small class. (Retention should not be used, unless in extreme cases). | STAR, Many studies of Retention (Holmes and Matthews). |
| IX. | Study costs and benefits; Use PTR and class size differences to get to small classes. | STAR, SSS, PTR studies Darling-Hammond; Miles |
| X. | Small classes and schools encourage increased student participation in schooling. | Finn, Voelkl, STAR, LBS, Lindsay's work, etc. |
| XI. | Small classes in early grades provide long-term multiple benefits. | Krueger; STAR Follow-up. Finn & Achilles, etc. |

* Detailed references are available. They were omitted from the table because of space and to improve the presentation.

# APPENDIX C

**68th LEGIS—2nd CALLED SESSION** **CH 28, SEC. IV-E1**

PART C. CLASS SIZE (P. 167)

**SECTION 1.** Section 16.054, Education Code, is amended to read as follows:

Sec. 16.054. STUDENT/TEACHER RATIOS; *CLASS SIZE. (a) Except as provided by Subsection (b) of this section, each* [~~Each~~] school district must employ a sufficient number of certified teachers to maintain an average ratio of not less than one teacher for each *20* [~~25~~] students in average daily attendance.

*(b) Beginning with the 1985-1986 school year, a school district may not enroll more than 22 students in a kindergarten, first or second grade class. Beginning with the 1988-1989 school year, a school district may nor enroll more than 22 students in a third or fourth grade class. This requirement shall not apply during the last 12 weeks of any school year.*

*(c) In determining the number of students to enroll in any class, a district shall consider the subject to be taught, the teaching methodology to be used, and any need for individual instruction.*

*(d) On application of a school district, the commissioner may except the district from the limits in Subsection (b) of this section if the commissioner finds the limits work an undue hardship on the district. An exception expires at the end of the semester for which it is granted, and the commissioner may nor grant an exception for more than one semester at a time.*

*(e) The commissioner shall report to the legislature each biennium regarding compliance with this section. The report must include:*

*(1) a statement of the number of school districts granted an exception under Subsection (d) of this section; and*

*(2) an estimate of the total cost incurred by school districts in that biennium in complying with this section.*

## APPENDIX D: One Major Caveat on Texas Data

Although the data presented in this paper up to this point, and the data from which the tables and conclusions are derived (from Texas sources) make a strong case that the smaller classes engendered by HB 72 (1984) are a major and fundamental reason for the apparent successes and achievement-gap reduction, there is a clear dissenting position. An alternative position was expressed recently by Haney (2000), an expert witness in Texas Assessment of Academic Skills (TAAS) litigation. The book-length presentation ("The Myth of the Texas Miracle in Education") attributed much of the TAAS test-score gains to three phenomena: retention in grade and its partner, dropout, and test-exempted students.

Haney accumulated massive data sets questioning the validity of the TAAS, and interpretation of TAAS results. For example, based on National Assessment of Education Progress (NAEP) data 1992-1998, Haney concluded ". . . the magnitudes of the gains apparent on NAEP for Texas fail to confirm the dramatic gains apparent on TAAS.", and "The Texas 'miracle' is more myth than real" (Part 7, p. 22). Haney (2000) used other data to question TAAS outcomes. For example, percentages of students taking the Texas Academic Skills Program (TASP) tests and passing all three parts (1992-93 to 1996-97) have shown a sharp and consistent decline, especially minority students (Adapted from Haney, 2000, Part 7, pp. 22-25).

Haney (pp. 22-23) reported taking the following information from the Texas Higher Education Coordinating Board web site that "describes the TASP testing program thus at its website" (www.thecb.state.tx.us/):

> . . . you can find out if you have the reading, writing, and math skills you need to do college-level work through the . . . TASP. The TASP Test, which is part of the TASP program is required—it is not optional.
>
> Beginning in fall 1998 you must take the TASP Test . . . before beginning classes at a public community college, public technical college, or public university in Texas . . .

If the grade-10 TAAS (reading, writing, and math skills) test results have been increasing dramatically, one would expect gains in TASP results on the same skills. Table D-1 shows a dramatic decline in TASP results while the TAAS rates supposedly climbed.

Table D-1. TASP Student Performance (1993-1997) Pass Rates by Race/Ethnicity

| | All Three Parts of TASP (% Rounded). | | | | | |
|---|---|---|---|---|---|---|
| YEAR | All | Native Am. | Asian | Af. Am. | Hisp. | White |
| 92-93 | 78 | 83 | 80 | 58 | 68 | 84 |
| 93-94 | 65 | 65 | 63 | 44 | 53 | 73 |
| 94-95 | 52 | 53 | 54 | 31 | 38 | 60 |
| 95-96 | 48 | 58 | 49 | 30 | 35 | 57 |
| 96-97 | 43 | 42 | 49 | 25 | 31 | 53 |

A possible cause of the increasing TAAS scores and decreasing TASP outcomes is the current concern among many educators that an undue emphasis on "high-stakes" testing (such as the TAAS exit test) requires educators to teach to the test, providing a restrictive and narrow curriculum. In discussing observations in classrooms during Tennessee's project STAR, a large class-size experiment, Evertson and Randolph (1998 as reported in Achilles 1999, p. 66) noted, "The curriculum in all classes . . . was tightly aligned to achievement of basic skills that could be readily measured by standardized and curriculum-driven criterion-referenced tests." This trend, currently being fueled by legislative mandates in other states (e.g., New York) seems to have three predictable outcomes: 1) high retention-in-grade rates, 2) high dropout rates, and 3) fewer of those who pass the basic-skills tests who now have the knowledge and skills to enter and succeed in higher education.

Because the "must pass" TAAS data are mostly from grade 10, what happens prior to grade 10 is important. Haney (2000, Part 5) reviewed grade retention and dropout data. Students retained in grade 9 do not take the grade-10 TAAS with their cohort, and if retained, they may drop out. Haney provided data on retention by grades (K-12) by ethnicity. Ranges in percentage retained, grades K-8, for White (.60%-4.40%), African-American (.90%-7.0%) and Hispanic (1.0%-6.6%) favor White students, but only moderately. The picture changes dramatically at grades 9 and 10. (1996-1997 data). (Part 5, p. 7).

Table D-2: Percentage Retained (1996-97) grades 9 and 10, Texas

| Grade | White | Af. Amer. | Hispanic | Total |
|---|---|---|---|---|
| 9 | 9.6 | 24.2 | 25.9 | 17.8 |
| 10 | 4.8 | 11.6 | 11.4 | 7.9 |

Regarding retention in grade and achievement-gap reduction, Haney (Part 5, pp. 3-5) noted the differential in retention for white and other students in Texas.

> . . . since 1992 progress from grade 9 to high school graduation has been stymied for Black and Hispanic students not after grade 10 when they first take the TAAS exit test, but in grade 9 before they take the test (p. 3) . . .. Also, in 1998-99 there were 25-30% fewer Black and Hispanic students enrolled in grade 10 than had been enrolled in grade 9 in 1997-98 (p. 4).
>
> . . . the fact that by the end of the 1990's 25-30% of Black and Hispanic students, as compared with only 10% of white students, were being retained to repeat grade 9 instead of being promoted to grade 10, makes it clear that the apparent diminution in the grade 10 racial gap in TAAS pass rates is in some measure an illusion. (p. 4. Emphasis added in each quote).

The link between grade retention and dropout is well known. Texas data show that "Grade 10 enrollments in 1996-97 were about 16% lower than expected for Black and Hispanic students, but only 8% for White students." "What happened to these missing students?" (Part 5, p. 9). "The grade 9 retention rates in Texas are far in excess of national trends." (Part 5, p. 9).

Those grade-9 retention rates, one might surmise, result in increased TAAS exit-test scores at grade 10. They may also contribute to another statistic that seems to go with higher TAAS scores and higher retention-in-grade. Reporting on National Center for Education Statistics (NCES) data for 1990-92, 1993-95, and 1996-98, Haney (Part 7, p. 3) concluded: "As can be seen for all three time periods, these data show Texas to have among the lowest rates of high school completion among the 50 states" (p. 3). The NCES data directly contradict the results put forth by the Texas Education Agency (TEA), but they are supported by data provided by the Intercultural Development Research Association (IDRA) and shown in Haney (Part 7) as Table 7-1 (p. 3). Haney also noted that increases in special education-classified TAAS test takers help explain the Texas "Miracle": "Special education" may "also account for a portion of

the closing of the "race gap" in TAAS scores over this period" (Part 5, pp. 10-11). He noted a change in the passing rates for the General Education Development (GED) test that could influence both students staying in high schools and student self-report of completing high school (via GED). (Haney, Part 7).

This Appendix explores explanations to the achievement-gap reduction evident in Texas data other than class size that was reduced by Texas law (HB 72) in 1984. The School Administrator (2000, December) carried an article claiming that achievement increases in Texas were likely influenced positively by the "states current accountability system, which was initiated in 1992 . . ." (Johnson, Treisman & Fuller, 2000, p. 20). They also acknowledged that there had been "other statewide initiative over the past decade that have influenced student performances. These include class-size reduction in the early grades, increased local control and state funding of extended-year programs and pre-kindergarten" (p. 20). Class size, extended year, and public pre-kindergarten all have strong research bases, something not claimed for testing efforts.

Based upon Haney's careful analyses, the claim that HB 72-generated small classes have contributed substantially to reducing the achievement gap between students of various ethnicities on TAAS results in Texas (1995-2000), may be premature. Additional analyses are required.

Alignment of the increasing passing rates on the TAAS beginning in 1995-96 with implementation of research-supported improvements (HB 72) in 1985-86 provides strong but circumstantial evidence that testing is not the cause. But, this will take further analyses, too. One thing seems certain given the increasing retention-in-grade, the decrease in TASP scores, and the dismal high school graduation rates: There will be a large pool of persons to labor in unskilled occupations in Texas . Researchers might review the economics of this.

References:

Haney, W. (2000, August 19). The myth of the Texas miracle in education. Education Policy Analysis Archives, 8 (41). ISSN 1068-2341.

Bobbett, G. C., French, R. L. & Achilles, C. M. (1996, March 9). How meaningful are Texas school district report cards?. Paper at the American Association of School Administrators (AASA) annual meeting, San Diego, CA.

Achilles, C. M. (1999). Let's Put Kids First, Finally. Thousand Oaks, Corwin Press.

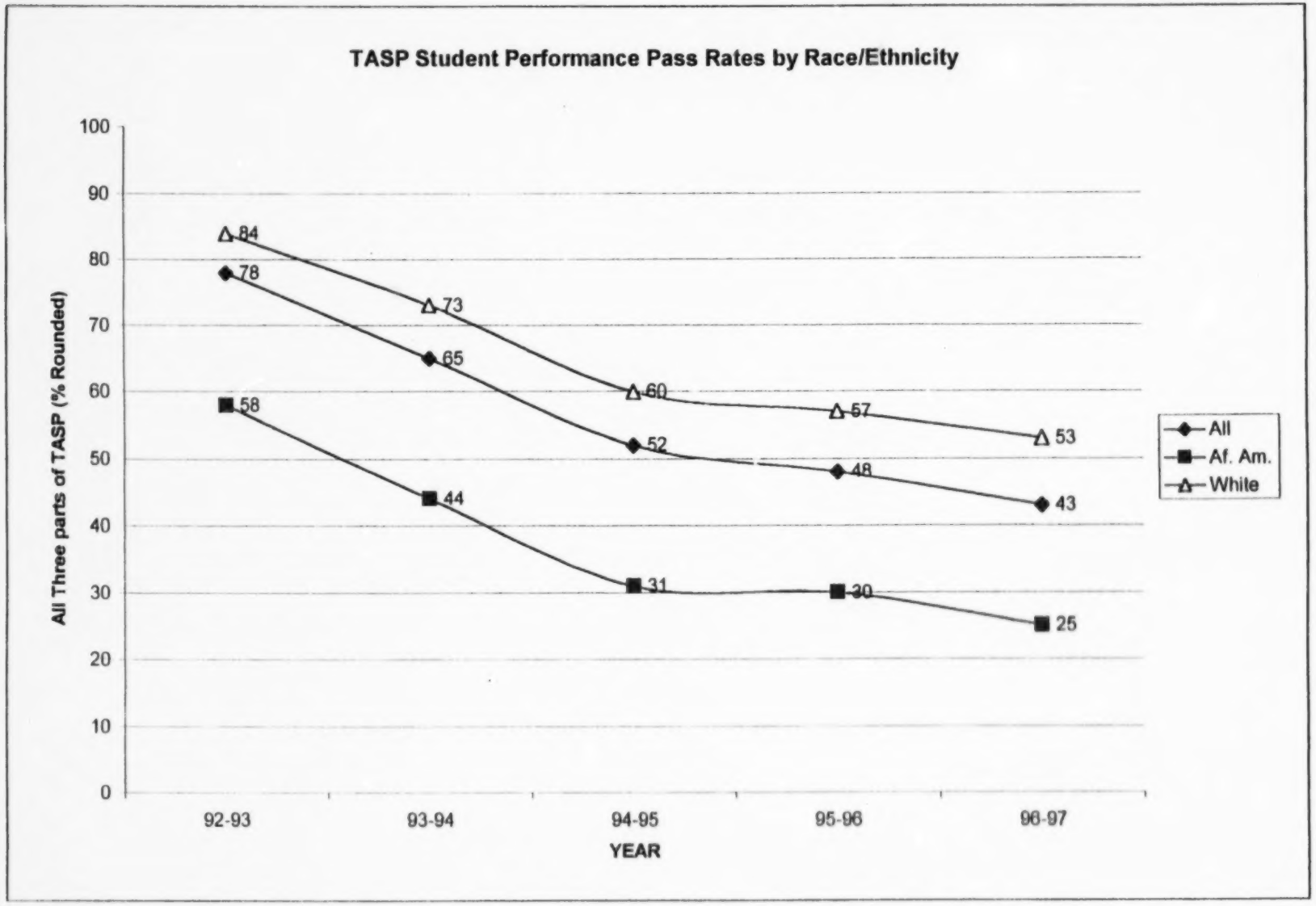



Figure D-1 TASP Student Performance (1993-1997) Pass Rates by Race/Ethnicity

APPENDIX E *

# THE BLACK-WHITE TEST SCORE GAP

Christopher Jencks, Harvard University; Meredith Phillips, UCLA

## THE BLACK-WHITE TEST SCORE GAP

- Reading and Mathematics scores rose for both African American and white students at all ages between 1971-1996, but African Americans increased more
- African American students scored lower than white students in vocabulary, reading, and mathematics tests, as well as tests that claim to measure scholastic aptitude and intelligence
- African American children arrive at kindergarten with few reading skills even when their parents have equal years of schooling
- Even though the gap has narrowed substantially, racial disparities are still very large

## RESEARCH FINDINGS

- Research indicates that ½ of the Black-White Gap in 12th grade reading and math was explained by differences in 1st grade scores
- The teacher expectations have more impact on African American students' performance than on white students' performance
- The test score gap between economically advantaged and disadvantaged children widens more over the summer then during the school year
- Exhorting teachers to have more faith in black children's' potential is unlikely to change their expectations
- Teachers expect less of African American students than white students because of student past performance and behavior (no evidence that teacher expectations differed by race)
- Rigorous interventions do effect IQ and cognitive skills at every stage of life
- Most of the divergence between African American students and white students with initially similar scores occur before high schools
- Evidence is mixed regarding whether African American children learn more in classes taught by African American teachers
- Class size matters more for African Americans
- Research tends to find that teachers who attend better colleges or scored higher on standardized examinations are more successful helping children score higher
- Different parenting practices may account for some of the black-white gap

## WHAT CAN SCHOOLS/SCHOOL SYSTEMS DO?

- Increase the number and quality of preschool programs (more focus on cognitive skills)
- Raise teacher expectations
- Increase academic preparation of teachers
- Provide staff development programs that focus on successful instructional methods/programs
- Reduce class size in the early years
- Promote better parenting practices/programs for all parents (match the school and home environments)
- Get support from community organizations to attack this gap problem
- Raise academic preparation of teachers
- Improve summer programs especially for at-risk students
- Implement programs that have been effective in other systems
- Encourage students to take more rigorous academic programs
- Success comes from coordinating human resources around a well defined goal, constantly assessing progress toward that goal and never giving up until success is achieved

* Summary provided by Michael E. Glascoe as a class assignment (12/00).